Александров Валентин, 620 группа

Отчет о заданию домашнему заданию 2 (от 14.09.20)

В качестве исходного литературного текста был выбран «Граф Монте-Кристо» Александра Дюма. Были обработаны первые 1000 строк текста, так как время обработки всего текста заняло бы приблизительно 18 часов. Обработка 1000 строк с помощью Mystem в два потока заняла 33 минуты. Из получившихся лемм убраны пустые леммы, леммы, содержащие небуквенные символы и русские стоп-слова.

Первые 100 лемм по частоте:

сказать              - 248

дантес               - 227

который              - 189

это                  - 146

вильфор              - 116

свой                 - 114

данглар              - 114

фернан               - 105

кадрусс              - 101

отвечать             - 99

знать                - 83

мочь                 - 83

мерседес             - 81

весь                 - 77

говорить             - 77

человек              - 71

эдмон                - 71

капитан              - 65

моррель              - 62

видеть               - 61

господин             - 61

рука                 - 61

отец                 - 57

спрашивать           - 54

письмо               - 53

дело                 - 50

хотеть               - 50

ваш                  - 41

глаз                 - 40

друг                 - 39

арматор              - 38

ничто                - 38

старик               - 38

де                   - 36

продолжать           - 36

идти                 - 35

наш                  - 35

самый                - 33

слово                - 32

маркиз               - 32

молодой              - 30

время                - 30

думать               - 29

взять                - 29

рене                 - 29

взгляд               - 28

голова               - 28

любить               - 28

казаться             - 28

жандарм              - 28

фараон               - 27

первый               - 26

корабль              - 26

становиться          - 26

выходить             - 26

лицо                 - 26

понимать             - 26

год                  - 25

сделать              - 25

голос                - 25

счастие              - 25

исполнять            - 24

пойти                - 24

прокурор             - 24

давать               - 23

место                - 23

должный              - 23

помощник             - 23

минута               - 23

оставаться           - 23

вскричать            - 23

вино                 - 23

марсель              - 22

улыбка               - 22

дверь                - 22

стол                 - 22

лодка                - 21

ждать                - 21

королевский          - 21

моряк                - 20

очень                - 20

иметь                - 20

умирать              - 19

день                 - 19

сын                  - 19

остров               - 18

отдавать             - 18

увидеть              - 18

оно                  - 18

невеста              - 18

каталанец            - 18

тюрьма               - 18

оставлять            - 17

узнавать             - 17

бумага               - 17

делать               - 17

забывать             - 17

сердце               - 17

твой                 - 17

стакан               - 17